Proceedings of the Asia-Pacific Advanced Network 2010 v. 30, p.10-13.
http://dx.doi.org/10.7125/APAN.30.2 ISSN 2227-3026

Information Hiding in Email Services Based on Confused Document Encrypting Schemes

Wei-Shyun Pan
Graduate Institute of Communication Engineering
National Chi Nan University, Nantou
54561, Taiwan
s97325527@ncnu.edu.tw

Quincy Wu
Graduate Institute of Communication Engineering
National Chi Nan University, Nantou
54561, Taiwan
solomon@ncnu.edu.tw

ABSTRACT
Confused Document Encrypting Scheme (CDES) is a document protection technique for information hiding (steganography), and it uses meaningful cheating text to confuse attackers. It is a simple and effective file protection scheme . In this paper, we proposed a new web application based on CDES, which apply with the email service, and we focus on the web service development and improve some known issues in CDES by a new method which is adding image-hiding technique to protect the position index file (PIF), and we choose a suitable compression and encryption algorithm to increase performance for CDES in my system.

Keywords
Information hiding (Steganography); Confused Document Encrypting Scheme(CDES) ;Email service; position index file(PIF);.

1. INTRODUCTION
As the rapid development of Internet and increasing popularity of personal computers, the behavior of human communication has been changed dramatically. Meanwhile, the security of Internet communication has also become the most popular topic for many researchers. To protect "personal privacy", two security techniques were proposed, which are cryptology and information hiding (steganography). Given a plaintext important message, cryptology utilizes an encryption algorithm and a secret key, to perform an encryption operation which converts the message to unreadable form for human. This encrypted cipher text can then be transmission safely on Internet. Theoretically, no one except the receiver, who owns the correct key, can decrypt the cipher text to obtain the original plaintext. This is the tradition way to protect important messages when they are transmitted on insecure channels like Internet.

However, encrypted messages usually looks meaningless, so if important messages are always encrypted, it is easy for attackers to notice that and they would be encouraged to allocate more resource to decrypt this encrypted messages. On the contrary, information hiding tried to hide information in some media, so that for anyone who intercepts the contents, it would just looks as usual messages that is irrelevant to the confidential information. Only the receiver who decrypts the messages with a proper key can successfully see the confidential information.

There are many good research papers which study how to hide information in an image file. However, it is still a difficult problem to hide information into a text message. Because the data volume of a text file is generally smaller (compared to an image file), any modification to the original text file will be easily noticed. To overcome this problem, a technique called Confused Document Encrypting Scheme (CDES) was proposed for information hiding.

In CDES, the basic idea is to hide important messages in cheating text which are exactly meaningful messages. The sender needs to transmit many cheating texts and an encrypted file. When an attacker intercepts these cheating texts, he will believe that these are normal texts and ignore them. In the following section, we briefly describe some variations of CDES and its disadvantages, then we propose a new web message service based on CDES and demonstrate how e-mail services can be protected in this framework.

2. RELATED WORK

2.1 Confused Document Encrypting Scheme
All Confused Document Encrypting Scheme (CDES) is a document protection scheme proposed in 1998 by Lin and Lee [1]. The basic concept is that it is unnecessary to directly send an encrypted plaintext. On the contrary, it will send many meaningful cheating texts and an encrypted Plaintext Index File (PIF) to confuse the attackers.

The algorithm of CDES cryptosystem is:

Input:

1. A plaintext file
2. Several cheating texts (The cheating text characters must at least include those of the plaintext characters).
3. Two secret keys with 128bits by International Data Encryption Algorithm (IDEA) encrypting algorithm. The first key is to encrypt the plaintext index file (PIF) and the second key is to encrypt the cheating text ID).

Output:

1. Several cheating texts with an IDs
2. Encrypted PIF

Step 1: Use the cheating text to generate the character position table (CPT), The table will store frequency and positions of characters in cheating text.

Step 2: Using the CPT and the Plaintext to generate a plaintext index file. By look up each character in sequence, and comparing with character elements of CPT, and choose a randomly value of position record to store in the PIF.

Step 3: Compress the PIF (You can use any compression algorithm).

Step 4: Randomly generate a IDs for each cheating texts.
Step 5: Use the first key to encrypt the compressed PIF.
Step 6: Use the second key to encrypt the correct cheating text's ID and put the encrypted ID in the head of the encrypted PIF.
Step 7: Send out many cheating texts with IDs and the encrypted PIF.

Example:
Input
Plaintext : Cat is my Pet.
{C, a , t , I , s , m , y , p , e , . , space}
Cheating text : Computer security is important.
{C, o , m , p, u , t , e , r , s , c , i , y , a , n , . ,space}
In Table 1, CPT is generated according to the above cheating text,.

Table 1. Characters Position Table(CPT)

Character	Position record
C	1
O	2 , 25
M	3 , 23
P	4 , 24
U	5 , 13
T	6 , 16 , 27 , 30
E	7 , 11
R	8 , 14, 26
S	10 , 20
C	12
I	15 , 19 , 22
Y	17
A	28
N	29
.	31
Space	9 , 18 , 21

And the CPT comparing with plaintext, according all of the characters of plaintext to choose a randomly position value from CPT , and then, we can store these values to the PIF.
The content of PIF is:
1 28 16 21 15 20 18 3 17 9 24 7 6 31
After the PIF was generated, assign a random ID for the cheating text
NO: 236785
Computer security is important.
Finally, after continually compression and encrypting processing, and it will send out the encrypted PIF with ID and all of the cheating texts to the receiver, and the receiver want to reverse original data, who needs use the ID with cheating text to generate CPT and use the position information in the PIF to find out the character, and then we will got the original plaintext.
According above mentioned, we can know the CDES is a very simple and effective technique for information hiding.

2.2 Further improvement

First proposed in CDES with Lin & Lee, it had several problems existed:

1. **Language support**

   In the CDES, the characters covers the range of 0 and 127, but the Chinese font (BIG5) had about 5401 characters, so it only support the English. Later, Yen & Hwang [2] use the characteristics of Chinese internal code, which translates the Chinese word to hexadecimal code , because it only needs to use the characters range of '0' to '9' and 'A' to 'F'. They were success to solve the problem.

2. **The cheating text must contain all characters in secret message**

   Based on the problem, Liang, et al. proposed a method [3], which automatically append the missing character of '0' to '9' and 'A' to 'F' of CPT generating in progress. So we do not care what content in the cheating text now.

3. **Large size of PIF**

   The PIF be to four to eight times as large as the original secret message. Therefore, Yao [4] proposed a method to solve it, and the PIF success to be two times compared with original data . and reduce the overhead in PIF generated of CDES.

2.3 Issues of cheating text transmitted

In previously proposed in [1] [2] [3],they separate trans-mission files in cheating texts and PIF, Because these behaviors are suspicious, and it will waste the bandwidth in transmission, but in [4], the receiver must use the encrypted URL (Uniform Resource Locator) to get cheating text.

3. Our proposed scheme

Integrates with previous problems of section II, we will propose an application in message service over internet based on Confused Document Encrypted Scheme.

First, the Fig.1 is my system model proposed, it composed of four modules, and we will describe these modules in sequence.

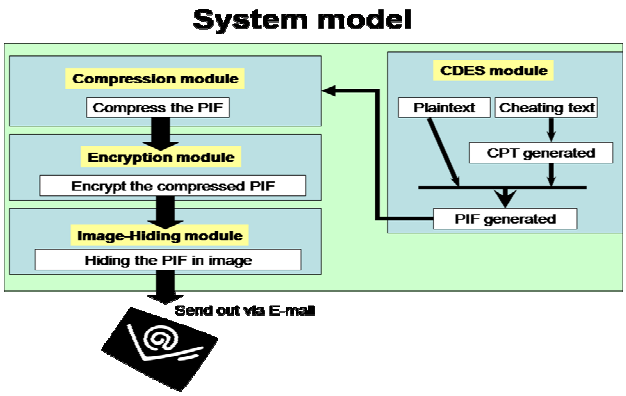

Figure 1. SYSTEM MODEL

3.1 CDES module

The module is based on the original file protection mechanism of CDES and adding other improved proposed in [2] [3] [4] for development. In my CDES, first, the CPT generated steps is not the same as original CDES, it will calculate the frequency all of characters ('0' to '9' and 'A' to 'F') and according the values of frequency to assign the range of index value. Second, I remove the randomly ID for cheating text.

3.2 Compression module

In the module, because the size of PIF is too large, so I need to use the LZMA algorithm to compress it, and the algorithm get

better performance with highly compression rate for my system, because it is a dictionary coding algorithm, therefore, it can reduce and compressing the same string in your files.

3.3 Encryption module

The module can provide encryption for the PIF, and I use the Blowfish algorithm, which is a symmetric block cipher algorithm, and the length of secret-key is 32 to 448 bits. The algorithm is a fast encrypting rate and high security encryption algorithm.

3.4 Image-Hiding module

It is a main method in my concept of my system, I use the image hiding technique to hide my PIF file in the photo, and it is an static or dynamic emoticon. For example, ☺ (Smile face), ☹ (Sad face).

3.5 Elements of CDES

- Plaintext: the secret message.
- Cheating text: The text is using for confuse the hackers.
- CPT(character position table) : Calculating and assign values to frequency count and range of index from characters of cheating text
- PIF (Plaintext index file): Using the CPT and Plaintext to choose a index value and record.
- Key: The first-key is using for PIF and the second-key is using for image.

3.6 Flow chart

Sender (Fig.2)

Step 1: Read the Plaintext

Step 2: Read the Cheating text

Step 3: Translates the contents of Plaintext and Cheating text to hexadecimal code.

Step 4: According the cheating text to generate CPT (note: The rule of CPT generated, it will limit the characters frequency must be smaller than or equal to 16, so the scope of index value will exceed 256.)

Step 5: Plaintext generates the PIF by CPT

Step 6: Compress the PIF

Step 7: Use the first-key to encrypt PIF

Step 8: Hiding the PIF in image (emoticon or any photo) and use the second-key to encrypt the image.

Step 9: Compose a mail to receiver (We must attaches the image file in mail.

Step 10: Send out the mail

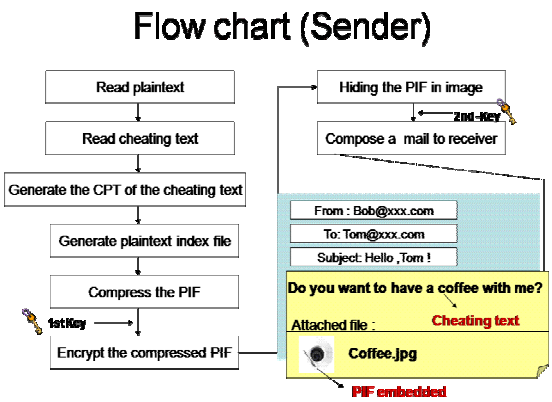

Figure 2. Flow chart in Sender

Receiver was described in Fig. 3.

Step 1: Receive a mail on any mail client

Step 2: Reading the cheating text from body of mail

Step 3: Using the second-key to decrypt and seeking the image to reveal the PIF from attached file of mail

Step 4: Using the first-key to decrypt the PIF

Step 5: Translates the cheating text to hexadecimal code

Step 6: Generating the CPT by cheating text

Step 7: Decompress the PIF

Step 8: Using the index position information of PIF and CPT to generate entire characters for original plaintext

Step 9: Plaintext Reversed and output to console

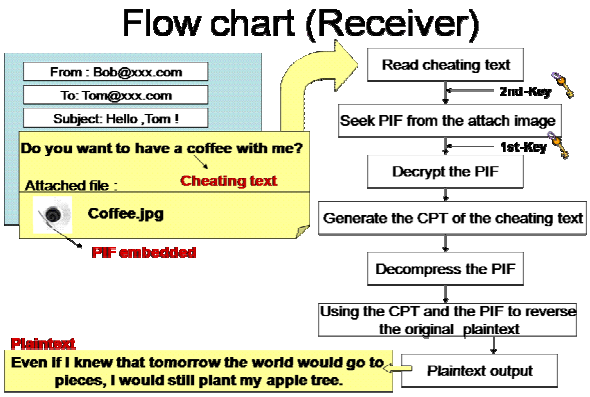

Figure 3. Flow chart in Receiver

4. Performance Evaluation

4.1 Environment

I am using the windows system and the IDE is visual studio 2005 and using the Mutt (Free mail client on Freebsd) to compose and send out my mail.

4.2 Performance on PIF compression

In table 2, we show the reason for LZMA algorithm chosen.

Table 2. Characters Position Table(CPT)

Plaintext (byte)	PIF size (byte)	Compression (byte)
241	1262	286
1227	2682	1000
Special case of LZMA		
635	3810	173

The special case mean the counts of duplicate string will effect the compression rate of LZMA

4.3 Demonstration

Input

Plaintext : Even if I knew that tomorrow the world would go to pieces, I would still plant my apple tree.

Cheating text : Do you want to have a coffee with me?

Image : coffee.jpg

1. Compose a mail by mutt

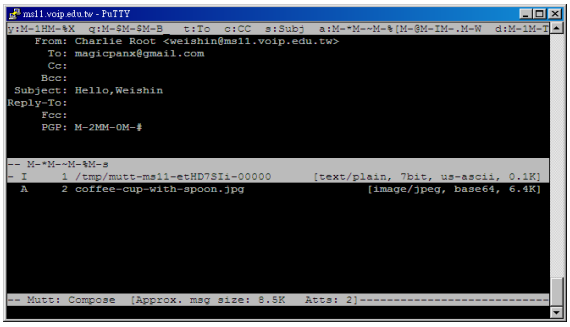

Figure 4. mail editor in mutt

2. Receive the mail

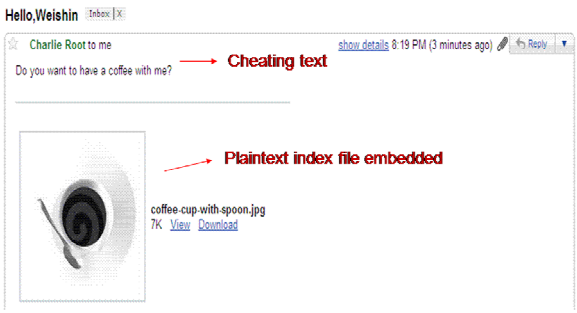

Figure 5. Receives the mail in mailbox

3. According cheating text and image to generate the plaintext, and then output these text to the console.

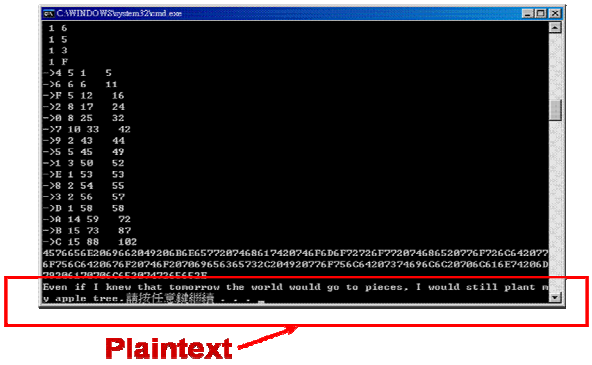

Figure 6. Show out the original plaintext

5. Conclusion and future work

In this study, we had observed the pattern of human in communication over internet, and the interesting thing became to our concept in the research.

We success to apply and improve the CDES for the email service, so we can see the system proposed will show some advantages, like increase the security and reduce the cost in the transmission, but the most important thing is we solve the problem of traditional CDES in Web application .

Moreover, the system can use for instant message (IM), like Live messenger, Yahoo talk, Google talk in the future work, because use the emoticon and photo in the chat with human , it has been a common behavior.

6. REFERENCES

[1] Chu-Hsing Lin and Tien-Chi Lee,"A Confused Document Encrypting Scheme and its Implementation", Computers & Security,Vol.17, No. 6, pp.543-551, 1998

[2] Wen-Hung Yeh and Jing-Jang Hwang, “Hiding Digital Information Using a Novel System Scheme”, Computers & Security, Vol. 20, No. 6, pp.533-538, 2001.

[3] Bi-feng Liang, etc, “On the study and implementation for confused document encrypting scheme of data hiding”, Technical Report, Department of InformationManagement, Ta Hwa Institute of Technology, R.O.C.,2002.

[4] Tzu-jung Yao and Quincy Wu, "On the Study of Overhead Reduction for Confused Document Encrypting Schemes", International Conference on Multimedia Computing and Information Technology (MCIT 2010) University of Sharjah(UoS), Sharjah, United Arab Emirates (UAE), March 2-4, 2010